The Effectiveness of the Linear Hull Effect

S. Murphy

Technical Report
RHUL–MA–2009–19
16 October 2009

Department of Mathematics
Royal Holloway, University of London
Egham, Surrey TW20 0EX, England

http://www.rhul.ac.uk/mathematics/techreports

Abstract

There is no linear hull effect in linear cryptanalysis.

1 Introduction

Linear cryptanalysis [7] is one of the standard techniques of assessing the security of block ciphers and is based on linear approximations to the plaintext, ciphertext and key. In our discussion of linear approximations, we consider an iterated block cipher encryption with plaintext p, ciphertext c and extended key k, where p, c and k are binary column vectors. We define the extended key k as the concatenation of the the round subkeys derived from the block cipher key through the key schedule. A *linear approximation* in its most basic form is usually regarded as a statement of the form

$$\alpha^T \begin{pmatrix} p \\ c \end{pmatrix} = \alpha_P^T p + \alpha_C^T c = \gamma^T k \text{ with probability } \tfrac{1}{2}(1+\epsilon_\gamma).$$

The vectors α, α_P and α_C are known respectively as the (overall) data mask, the plaintext mask and the ciphertext mask, with $\alpha^T = \left(\alpha_P^T\ \alpha_C^T\right)$, and the vector γ as the key mask. The value ϵ_γ is known as the *imbalance* or *correlation* of the linear approximation, and $\frac{1}{2}\epsilon_\gamma$ as the *bias* of the linear approximation. If the linear expression is *unbiased*, that is $\epsilon_\gamma \neq 0$, then the linear expression can potentially be used to give an estimate of one bit $\gamma^T k$ of key information. The number of plaintext-ciphertext pairs required to estimate this key bit to a required accuracy is proportional to ϵ_γ^{-2}. This is the procedure given by Algorithm 1 of [7].

The more sophisticated Algorithm 2 of [7] uses trial encryptions and decryptions of the outer rounds under various partial subkeys. Under certain assumptions, the technique of Algorithm 2 is equivalent to constructing a method of distinguishing the distribution of the linear expression

$$\alpha^T \begin{pmatrix} p \\ c \end{pmatrix} + \gamma^T k,$$

which is zero with probability $\frac{1}{2}(1+\epsilon_\gamma)$, from a uniform distribution [7, 8]. As before, the number of plaintext-ciphertext pairs required to make this distinction to a required accuracy is proportional to ϵ_γ^{-2} [7].

The usual method of calculating the probability that such a single linear expression holds is to use the so-called *Piling-Up Lemma* of [7]. However, in the analysis of many block ciphers, such use of the *Piling-Up Lemma* can generate a number of linear expressions with the same data mask but differing key masks, that is a key mask set Γ and a collection of expressions of the form

$$\alpha^T \begin{pmatrix} p \\ c \end{pmatrix} = \gamma^T k \text{ with probability } \tfrac{1}{2}(1+\epsilon_\gamma) \quad [\gamma \in \Gamma].$$

The existence of a number of such "unbiased" expressions ($\epsilon_\gamma \neq \frac{1}{2}$) generated by the *Piling-Up Lemma* appears to be the motivation for the concept of the *linear*

hull of linear expressions introduced by [8]. The *linear hull* for data mask α is the set of all such above expressions for different key masks γ. It is asserted by [8] that existence of such a linear hull containing many such unbiased expressions generally increases the efficiency of Algorithm 2 of [7]. Furthermore, such an assertion appears to be generally accepted and widely used in the analysis of block ciphers. For example, a standard reference work on cryptology makes the following statement about the use of the linear hull in cryptanalysis [3].

> LINEAR HULLS: Estimating the bias of approximations by constructing linear characteristics is very convenient, but in some cases, the value derived in this way diverges significantly from the actual bias. The most important cause for this difference is the so-called *linear hull* effect, first described by Nyberg in 1994 [8]. The effect takes place when the correlation between plaintext and ciphertext bits, described by a specific linear approximation, can be explained by multiple linear characteristics, each with a non-negligible bias, and each involving a different set of key bits. Such a set of linear characteristics with identical input and output masks is called a *linear hull*. Depending on the value of the key, the different characteristics will interfere constructively or destructively, or even cancel out completely. If the set of keys used in different linear characteristics are independent, then this effect might considerably reduce the average bias of [a single linear] expression, and thus the success rate of the simple attack described above [Algorithm 1 of [7]]. Nyberg's paper [8] shows, however, that the more efficient attacks [Algorithm 2] described in [7], which only use the linear approximations as a distinguisher, will typically benefit from the linear hull effect.

We show that this so-called *linear hull effect* [3, 8] simply does not exist.

2 The Fundamental Probability

Our analysis of the linear hull effect is based on the fundamental probability (Definition 1), together with the related concept of the fundamental imbalance (Definition 2). The fundamental probability is a well-defined probability for use in linear cryptanalysis, particularly in analysing the linear hull effect.

Definition 1. The *fundamental probability* $q(k)$ of data mask α for block cipher encryptions under the fixed extended key k is

$$q(k) = \mathbf{P}_k \left(\alpha^T \begin{pmatrix} p \\ c \end{pmatrix} = 0 \right)$$

Definition 2. The *fundamental imbalance* of data mask α under fixed key k with fundamental probability $q(k) = \frac{1}{2}\left(1 + \eta(k)\right)$ is

$$\eta(k) = \mathbf{P}_k \left(\alpha^T \begin{pmatrix} p \\ c \end{pmatrix} = 0 \right) - \mathbf{P}_k \left(\alpha^T \begin{pmatrix} p \\ c \end{pmatrix} = 1 \right) = 2q(k) - 1.$$

3 Probabilistic Interpretation of the Linear Hull

We noted that the fundamental probability (Definition 1) is well-defined. However, the probability statements made in the usual definition of a linear hull are not in general well-defined. To demonstrate this, we consider very carefully the nature of the statements about linear expressions used in the linear hull for a data mask α.

In the standard formulation of linear cryptanalysis given in Section 1, a linear approximation is defined directly in terms of a key mask γ, that is a linear approximation is a statement of the form

$$\alpha^T \begin{pmatrix} p \\ c \end{pmatrix} = \gamma^T k \text{ with probability } \tfrac{1}{2}(1+\epsilon_\gamma),$$

which only depends on the key through the value of $\gamma^T k$. In terms of the fundamental probability for data mask α, we have

$$\mathbf{P}_k\left(\alpha^T \begin{pmatrix} p \\ c \end{pmatrix} = 0\right) = \tfrac{1}{2}(1+\eta(k)) = \begin{cases} \frac{1}{2}(1+\epsilon_\gamma) & [\gamma^T k = 0] \\ \frac{1}{2}(1-\epsilon_\gamma) & [\gamma^T k = 1]. \end{cases}$$

Thus the fundamental imbalance $\eta(k)$ is given in terms of the usual form of the imbalance ϵ_γ used in linear cryptanalysis by $\eta(k) = (-1)^{\gamma^T k}\epsilon_\gamma$. However, for a fixed key k, the fundamental imbalance $\eta(k)$ is clearly a constant. Thus for the above probability to be well-defined, we require $(-1)^{\gamma^T k}\epsilon_\gamma$ to be constant for all non-trivial key masks γ for fixed k. We now discuss this issue in more detail.

The linear hull is usually considered to be defined by several probabilistic statements such as (for $\gamma \neq \gamma'$)

$$\begin{aligned} \alpha^T \begin{pmatrix} p \\ c \end{pmatrix} &= \gamma^T k \quad \text{with probability} \quad \tfrac{1}{2}(1+\epsilon_\gamma) \\ \text{and} \quad \alpha^T \begin{pmatrix} p \\ c \end{pmatrix} &= \gamma'^T k \quad \text{with probability} \quad \tfrac{1}{2}(1+\epsilon_{\gamma'}). \end{aligned}$$

According to these linear expressions, for a fixed key k, $\alpha^T \begin{pmatrix} p \\ c \end{pmatrix}$ has to simultaneously take the value $\gamma^T k$ with probability $\frac{1}{2}(1+\epsilon_\gamma)$ and the value $\gamma'^T k$ with probability $\frac{1}{2}(1+\epsilon_{\gamma'})$. This means there are four cases required to evaluate the fundamental probability $\mathbf{P}_k\left(\alpha^T \begin{pmatrix} p \\ c \end{pmatrix} = 0\right)$, as given in the following Table.

	$\gamma'^T k = 0$	$\gamma'^T k = 1$
$\gamma^T k = 0$	$\frac{1}{2}(1+\epsilon_\gamma) = \frac{1}{2}(1+\epsilon_{\gamma'})$	$\frac{1}{2}(1+\epsilon_\gamma) = \frac{1}{2}(1-\epsilon_{\gamma'})$
$\gamma^T k = 1$	$\frac{1}{2}(1-\epsilon_\gamma) = \frac{1}{2}(1+\epsilon_{\gamma'})$	$\frac{1}{2}(1-\epsilon_\gamma) = \frac{1}{2}(1-\epsilon_{\gamma'})$

$$\mathbf{P}_k\left(\alpha^T \begin{pmatrix} p \\ c \end{pmatrix} = 0\right)$$

We can therefore deduce that $\epsilon_\gamma = \epsilon_{\gamma'} = 0$. Thus the probability statements used in specifying a linear hull are not well-defined if there is more than one unbiased linear approximation in the linear hull. A linear hull in the sense of [3, 8], that is a collection of several linear expressions each with a significant imbalance or bias, is not a well-defined probabilistic concept.

4 Analogues in Differential Cryptanalysis of Linear Hulls

A much-repeated heuristic justification for the linear hull effect is given in [8] by claiming that the use of the linear hull is analogous to the use of the *differential* [6] in differential cryptanalysis [1, 2]. However, a detailed examination shows that this analogy is not sustainable. We therefore state the claim of [8] of an analogy between differential and linear cryptanalysis using our notation.

> We conclude that Algorithm 2 [of [7]] makes in fact use of a family of linear approximate expressions
>
> $$\alpha_P p + \alpha_C^T c + \gamma^T k = \alpha^T \begin{pmatrix} p \\ c \end{pmatrix} + \gamma^T k$$
>
> where [the data mask] α is fixed but [the key mask] γ varies. This means that the round approximations which uniquely determine γ and are uniquely determined by γ, can be chosen in all possible ways to form a chain of approximations from $\alpha_P^T p$ to $\alpha_C^T c$. Hence there is a close analog with what is called differentials in differential cryptanalysis [6].

The key mask γ does indeed determine a series of round data masks with the outer round masks constrained by α. These data masks give a random process [4, 5] with a state space consisting of two elements $\{0, 1\}$, in for example the manner described by [9]. The usual method of analysing this random process in linear cryptanalysis is to use the *Piling-Up Lemma* [7], which is applicable to this random process if it is a Markov process [9]. However, whether or not the *Piling-Up Lemma* is applicable, both states are considered by the *Piling-Up* calculation. Loosely speaking, the *Piling-Up Lemma* probability calculation may incorrectly "assign probability" to the states if the Markov assumption is not valid, but "all probability is assigned". There are no "ignored" states in linear cryptanalysis, and there is certainly no "missing probability" not considered by the *Piling-Up Lemma* waiting to be found.

In differential cryptanalysis, a pair of plaintexts is encrypted under the same fixed key. A random process is derived by taking the difference of those plaintexts at every rounds, so giving a random process with a state space of size 2^n. Under the assumption that this differential cryptanalysis random process is a Markov process, the *differential* [6] is calculated by using the product of the matrices of one-step (round) transition probabilities. The *characteristic* [1, 2] is essentially obtained by setting all but one element to 0 in each of these one-step transition matrices and then calculating the product of these revised matrices. A value for

a probability calculated using a *differential* [6] is therefore always at least that of the same value for a probability calculated using a *characteristic* [1, 2]. Thus there is a meaningful sense in differential cryptanalysis in which the differential can be thought of as "finding probabilities" missed by the characteristic, as the characteristic "ignores" most of the states in the random process.

The correct analogue of the linear hull in differential cryptanalysis is the *enhanced characteristic*, considered in Sections 5.1 and 6.5 of [1]. The enhanced characteristic gives an enhanced differential cryptanalysis random process in which the state at any round is the difference between the between the data values at that round (as for standard differential cryptanalysis) and also a collection of data values at certain bit positions. Similarly, in linear cryptanalysis, the simultaneous use of two key masks with the same data mask, really defines sequences of 2-dimensional data masks for the inner rounds. This gives rise to a random process with, in general, a 4-element state space rather than the 2-element state space given by a single key mask. For both the enhanced characteristic in differential cryptanalysis and the linear hull in linear cryptanalysis, the state space of the underlying random process state space is obtained by refining the standard differential cryptanalysis or linear cryptanalysis state space.

The linear hull therefore defines a new (enhanced) random process beyond the standard linear cryptanalysis random process. By contrast, probability calculations in differential cryptanalysis using either characteristics or differentials are conducted with respect to the same standard differential cryptanalysis random process. Thus a calculation using a linear hull and a calculation using a differential are fundamentally different in relation to their underlying standard random process. A differential can find "unused probability", whereas a linear hull simply cannot. There is no analogue between linear hulls and differentials in the manner claimed by [8].

5 Data Requirements for the Linear Hull Effect

The linear hull effect supposedly reduces the number of plaintext-ciphertext pairs required to distinguish to a given accuracy a given distribution from the uniform distribution [3, 8]. Before analysing this claim, we first calculate the data requirements exactly. For a fixed extended key k, the number N_k of plaintext-ciphertext pairs required to use the distinguisher to a required accuracy is asymptotically inversely proportional to $\left(q(k) - \frac{1}{2}\right)^2$ [7]. Under the assumption that all extended keys are equally likely, then the mean number (over all keys) N of plaintext-ciphertext pairs required to use the distinguisher to a specified degree of accuracy is given by $N = \mathbf{E}[N_k]$, for some suitably-defined expectation $\mathbf{E}$. If we let $\mathcal{K}$ denote the set of extended keys, then the mean number N of plaintext-ciphertext pairs required is proportional to

$$\mathbf{E}\left[\left(\left(q(k) - \tfrac{1}{2}\right)^2\right)^{-1}\right] = \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}\left(\tfrac{1}{4}\,\eta(k)^2\right)^{-1}.$$

We now consider the data requirement for the use of the linear hull given by [8]. The supposed existence of a linear hull effect depends on the assertion that this data requirement can be expressed in terms of a quantity defined by the *Fundamental Theorem* of [8]. We now examine this assertion. Accordingly, we let

$$r(\gamma, k) = \tfrac{1}{2}\left(1 + (-1)^{\gamma^T k}\eta(k)\right),$$

so $r(\gamma, k)$ is the quantity referred to as "$p(a, b, c; k)$" in [8], where "a, b" refer to the data mask α and "c" refers to the key mask γ. We note that "$p(a, b, c; k)$" is referred to as a probability by [8], but is not in general a well-defined probability (Section 3). The *Fundamental Theorem* considers the quantity ψ given by

$$\psi = \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}\left(r(\gamma, k) - \tfrac{1}{2}\right)^2 = \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}\tfrac{1}{4}\,\eta(k)^2 = \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}\tfrac{1}{4}\,\left(q(k) - \tfrac{1}{2}\right)^2.$$

Thus the quantity ψ considered by the *Fundamental Theorem* can be expressed in terms of the above expectation $\mathbf{E}$ as

$$\psi = \mathbf{E}\left[\left(q(k) - \tfrac{1}{2}\right)^2\right] = \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}\tfrac{1}{4}\,\eta(k)^2.$$

The claim for the reduction in data complexity given by the linear hull effect is then based on the assertion that the number N of plaintext-ciphertext pairs required to distinguish the distribution from uniform is proportional to ψ^{-1}, that is to say proportional to

$$\psi^{-1} = \mathbf{E}\left[\left(q(k) - \tfrac{1}{2}\right)^2\right]^{-1} = \left(\frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}\tfrac{1}{4}\,\eta(k)^2\right)^{-1}.$$

We now compare these two quantities for distinguishing the specified distribution from a uniform distribution. The actual expected number N of plaintext-ciphertext pairs and the number given by the application of the *Fundamental Theorem* in the manner of [8] are respectively proportional to

$$\mathbf{E}\left[\left(\left(q(k) - \tfrac{1}{2}\right)^2\right)^{-1}\right] \text{ and } \mathbf{E}\left[\left(q(k) - \tfrac{1}{2}\right)^2\right]^{-1}.$$

These two quantities can easily be compared by using Jensen's inequality [10], which we state in Lemma 1.

Lemma 1. (*Jensen's Inequality.*) A random variable X and convex function ζ satisfy $\zeta\left(\mathbf{E}[X]\right) \leq \mathbf{E}\left[\zeta(X)\right]$.

Inversion of the positive real numbers is a convex function, so Jensen's inequality gives

$$\mathbf{E}\left[\left(\left(q(k) - \tfrac{1}{2}\right)^2\right)^{-1}\right] \geq \mathbf{E}\left[\left(q(k) - \tfrac{1}{2}\right)^2\right]^{-1}.$$

The true data requirement is proportional to the left-hand side of the above inequality, whereas the quantity ψ considered by the *Fundamental Theorem* only addresses the right-hand side of this inequality. Thus the *Fundamental Theorem* can only ever be used to give a lower bound on the data complexity. We illustrate this point in Examples 1 and 2.

Example 1. We consider a linear cryptanalysis for data mask α where the fundamental probability is given by

$$q(k) = \mathbf{P}_k \left(\alpha^T \begin{pmatrix} p \\ c \end{pmatrix} = 0 \right) = \frac{1}{2} \left(1 + (-1)^{\gamma^T k} \epsilon_\gamma + (-1)^{\gamma'^T k} \epsilon_{\gamma'} \right),$$

where $\epsilon_\gamma, \epsilon_{\gamma'} \neq 0$. Thus the fundamental imbalance is given by

$$\eta(k) = (-1)^{\gamma^T k} \epsilon_\gamma + (-1)^{\gamma'^T k} \epsilon_{\gamma'}.$$

The true data requirement to distinguish this distribution from a uniform distribution to a given degree of accuracy is proportional to

$$\mathbf{E}\left[\left(\left(q(k) - \tfrac{1}{2} \right)^2 \right)^{-1} \right] = \frac{1}{4} \left(\frac{2}{(\epsilon_\gamma + \epsilon_{\gamma'})^2} + \frac{2}{(\epsilon_\gamma - \epsilon_{\gamma'})^2} \right) = \frac{\epsilon_\gamma^2 + \epsilon_{\gamma'}^2}{\left(\epsilon_\gamma^2 - \epsilon_{\gamma'}^2 \right)^2}.$$

By contrast, the approach of [8] based on the *Fundamental Theorem* asserts that the data requirement is proportional to

$$\mathbf{E}\left[\left(q(k) - \tfrac{1}{2} \right)^2 \right]^{-1} = \frac{1}{4} \left(2(\epsilon_\gamma + \epsilon_{\gamma'})^2 + 2(\epsilon_\gamma - \epsilon_{\gamma'})^2 \right)^{-1} = \frac{1}{\epsilon_\gamma^2 + \epsilon_{\gamma'}^2}.$$

However, $\left(\epsilon_\gamma^2 + \epsilon_{\gamma'}^2 \right) \left(\epsilon_\gamma^2 - \epsilon_{\gamma'}^2 \right)^{-2} > \left(\epsilon_\gamma^2 + \epsilon_{\gamma'}^2 \right)^{-1}$, as $\left(\epsilon_\gamma^2 + \epsilon_{\gamma'}^2 \right)^2 > \left(\epsilon_\gamma^2 - \epsilon_{\gamma'}^2 \right)^2$, so can see that

$$\mathbf{E}\left[\left(\left(q(k) - \tfrac{1}{2} \right)^2 \right)^{-1} \right] > \mathbf{E}\left[\left(q(k) - \tfrac{1}{2} \right)^2 \right]^{-1}.$$

For this example, we can easily see that the data requirement averaged over all extended keys is always larger than that asserted by the supposed linear hull effect.

Example 2. We consider the cryptanalysis of Example 1 for a block cipher where $\epsilon_{\gamma'} = \epsilon_\gamma$. Such a block cipher would be considered to possess a large linear hull effect in the "linear hull literature". The fundamental probability $q(k)$ for data mask α is given by

$$q(k) = \begin{cases} \frac{1}{2} \left(1 + (\epsilon_\gamma + \epsilon_{\gamma'}) \right. & [\gamma^T k = \gamma'^T k = 0] \\ \frac{1}{2} \left(1 - (\epsilon_\gamma + \epsilon_{\gamma'}) \right. & [\gamma_T k = \gamma'^T k = 1] \\ \frac{1}{2} & [\gamma^T k \neq \gamma'^T k]. \end{cases}$$

Clearly, for any key $k \in \ker(\gamma + \gamma')^T$, that is for half of all the extended keys, the distribution in question is uniform, so by definition is indistinguishable from

a uniform distribution. This is reflected in that fact that the true data requirement, proportional to $\left(\epsilon_\gamma^2 + \epsilon_{\gamma'}^2\right)\left(\epsilon_\gamma^2 - \epsilon_{\gamma'}^2\right)^{-2}$, is formally infinite. By contrast, the supposed linear hull effect gives the data requirement as being proportional to $\left(\epsilon_\gamma^2 + \epsilon_{\gamma'}^2\right)^{-1}$, a finite quantity.

For this example, the supposed linear hull effect is effectively asserting that it is possible to distinguish an indistinguishable distribution given by one key because there exists a distinguishable distribution given by some other key.

6 Conclusions

The usual method of quantifying the supposed linear hull effect assumes that expectation and inversion are operations on a random variable which commute. Jensen's inequality shows that this assumption is generally incorrect. Thus the supposed linear hull effect simply ignores Jensen's inequality, a fundamental result in probability theory and statistical inference. Furthermore, Jensen's inequality shows that the *Fundamental Theorem* of [8] can only ever give a lower bound for the data requirement for using a collection of "linear approximations" with the same data mask.

The linear hull effect, in the usual sense [3, 8] of always improving the average efficiency of Algorithm 2 of [7], is an illusion.

References

1. E. Biham and A. Shamir. Differential Cryptanalysis of the DES-like Cryptosystems. *Journal of Cryptology*, 4:3–72, 1991.
2. E. Biham and A. Shamir. *Differential Cryptanalysis of the Data Encryption Standard*. Springer–Verlag, 1993.
3. A. Biryukov and C. De Cannière. Linear Cryptanalysis for Block Ciphers. In H.C. Van Tilborg, editor, *Encyclopedia of Cryptography and Security*, pages 351–354. Springer, 2005.
4. D.R. Cox and H.D. Miller. *The Theory of Stochastic Processes*. Chapman and Hall, 1965.
5. G.R. Grimmett and D.R. Stirzaker. *Probability and Random Processes*. Oxford University Press, 2001.
6. X. Lai, J.L. Massey, and S. Murphy. Markov Ciphers and Differential Cryptanalysis. In D.W. Davies, editor, *Advances in Cryptology – EUROCRYPT 1991*, volume 547 of *LNCS*, pages 17–38. Springer–Verlag, 1991.
7. M. Matsui. Linear Cryptanalysis for the DES Cipher. In T. Helleseth, editor, *Advances in Cryptology – EUROCRYPT 1993*, volume 765 of *LNCS*, pages 386–397. Springer–Verlag, 1993.
8. K. Nyberg. Linear Approximation of Block Ciphers. In A. De Santis, editor, *Advances in Cryptology – EUROCRYPT '94*, volume 950 of *LNCS*, pages 439–444. Springer–Verlag, 1995.
9. S. Murphy and F. Piper and M. Walker and P. Wild. Maximum Likelihood Estimation for Block Cipher Keys. *Technical Report RHUL-MA-2006-3*, Royal Holloway (University of London), 1994. `http://www.ma.rhul.ac.uk/techreports`.
10. S.D. Silvey. *Statistical Analysis*. Chapman and Hall, 1975.